Statistische Hefte 27, 207–225 (1986)

Statistische Hefte
Statistical Papers

Regression analysis for discrete event history or failure time data

Alfred Hamerle

Received: April 17, 86; Revised version: September 4, 86

The paper deals with discrete-time regression models to analyze multistate - multiepisode models for event history data or failure time data collected in follow-up studies, retrospective studies, or longitudinal panels. The models are applicable if the events are not dated exactly but only a time interval is recorded. The models include individual specific parameters to account for unobserved heterogeneity. The explanatory variables may be time-varying and random with distributions depending on the observed history of the process. Different estimation procedures are considered: Estimation of structural as well as individual specific parameters by maximization of a joint likelihood function, estimation of the structural parameters by maximization of a conditional likelihood function conditioning on a set of sufficient statistics for the individual specific parameters, and estimation of the structural parameters by maximization of a marginal likelihood function assuming that the individual specific parameters follow a distribution. The advantages and limitations of the different approaches are discussed.

1. Introduction

Event history or failure time data are collected in follow-up studies, retrospective studies, and sometimes in longitudinal panels. The data record qualitative changes over time in some important variables. The main purpose of the statistical analysis of such event histories or failure times is to investigate the time it takes before a certain event occurs. Examples are job changes, changes in residence, lay-offs, births, marriages, divorces, deaths, etc. In addition, it is important to evaluate the association of exposure, treatment and prognostic factors with the distribution of time until the event occurs.

Sometimes there is only one episode or spell for each individual measuring the time interval between an initial event and a termination event. This applies in particular to survival analysis where the detection of a disease is the initial event and the patient's death is the termination event. In other applications of these methods individuals can experience repeatable events or failures and moreover, these events or failures may be of various kinds. This leads to general multiepisode - multistate models and the successive episodes represent durations in different states.

For the statistical analysis of such dynamic processes hazard rate models can be used where the hazard rate depends on independent variables. The statistical theory of duration data using hazard rate models is described by Kalbfleisch and Prentice (1980), Lawless (1982), Cox and Oakes (1984), Tuma and Hannan (1984), Blossfeld, Hamerle and Mayer (1986) and others. Hamerle (1984) surveys applications of duration models in different areas and considers a general approach in the multistate - multiepisode case. Hujer and Schneider (1986) investigate the data of the first wave of the Socioeconomic Panel and compare several hazard rate specifications.

Most of the methods for analyzing event history or failure time data assume that time is measured as a continuous variable. The analysis presented here is specifically intended for situations in which the time scale is genuinely discrete or in which there is substantial grouping of the response times into class intervals. The methods are applicable when the data are not available as essentially exact response times but when the data record only the particular interval of time in which each event or failure occurs. This applies in particular to longitudinal panels where event histories e.g. about employment status and other important qualitative changes between the successive panel waves are registered retrospectively in fixed-length periods. One of the new panel studies of this kind is the Socioeconomic Panel of the 'Sonderforschungsbereich 3" (see Hanefeld (1984)). Other applications are in medical work when patients are followed up and detailed information on each patient is collected at fixed intervals, or in sociological research when attention is given to qualitative changes that occur in specific time intervals.

If there are only a few time intervals or if the time units are large then many failures are reported at the same time and the number of ties becomes high. Then, strictly speaking, continuous-time techniques are inappropriate (Cox and Oakes, 1984, p. 99/100). Some continuous-time methods, especially the partial likelihood estimation procedure for Cox's Proportional Hazards model (see e.g. Kalbfleisch and Prentice, 1980, ch. 4) make use of the temporal order in which the failures or events occur and they cannot be applied directly when the data include tied observations. In the presence of ties an approximate partial likelihood function is widely used (Breslow, 1974). But when the number of ties becomes high, this approximation yields severely biased estimates (Cox and Oakes, 1984, p.103, Kalbfleisch and Prentice, 1980, p. 74/75). Kalbfleisch and Prentice (1980), p.75, emphasize that there is some asymptotic bias in both the estimation of the regression coefficients and in the estimation of its covariance matrix. This applies not only to the Cox model but also to fully parametrized specifications of the hazard rate. Moreover, the papers which deal with the derivation of the asymptotic properties of the estimators in hazard rate models (see e.g. Andersen and Gill (1982), Borgan (1984)) assume that ties only occur with zero probability.

In such situations discrete-time models are more suited for the analysis of failure time data. Several authors, including Thompson (1977), Prentice and Gloeckler (1978), Mantel and Hankey (1978), Allison (1982), Aranda-Ordaz (1983), Laird and Olivier (1981), Hamerle (1985), have studied discrete-time regression models for failure time data. Here we present a different approach and consider the general case where individuals can experience many events or failures as time goes on and the events or failures may be of multiple kinds. In addition to the failure times and types of failures of an individual some concomittant information on explanatory variables or prognostic factors is included in the model to study the relationships between these variables and failures. The covariates may be time dependent and random with distributions depending on the observed experimental history. The covariate process may include fixed and 'external' time dependent as well as 'internal' time dependent covariates (see Kalbfleisch and Prentice, 1980, ch.5.2). The models also contain individual specific parameters. The role of the individual specific components is to control for unobserved heterogeneity, e.g. for omitted variables. The regression coefficients which are common to all of the individuals in the sample may be time varying, i.e. they may depend on the time intervals. Here we present several discrete-time hazard rate regression models and discuss their advantages and limitations. We derive unconditional, conditional and marginal estimation procedures where our concern is with the regression parameters β_t, the structural parameters of the model.

We refer to related work by Heckman (1981a, 1981b), Chamberlain (1980,1985), Arjas (1984) and Arjas and Haara (1986). Heckman and Chamberlain consider discrete-time models for state probabilities in analyzing traditional panel data. With the exception of Chamberlain (1980) these studies only investigate the case of two states. The models are appropriate for typical panel studies where the individual's state is determined at particular points in time and where information about events between these successive points in time is not available. Here we are modeling transition probabilities which is more general if the state space contains more than two elements. But appropriate data is needed. 'Panel mortality' can also be incorporated. Our approach differs from the one considered by Arjas and Haara in that they use a binary logistic model and do not include individual specific parameters in their model. They derive asymptotic results for the estimated regression coefficients as the number of time intervals tends to infinity. Here we assume that observation time is finite and that there is a reasonable number of study subjects. Asymptotic properties here always concern the case where the number of study subjects tends to infinity.

2. A general discrete-time hazard rate model

Choosing some convenient point in real time as the origin, we split the time axis into successive intervals $t=1,2,\ldots$ The last time interval of the observation period is denoted by T and we consider probability models in discrete time $t=1,\ldots,T$.

The individuals included in the study are indexed by i, $i\geq 1$. It is not necessary for all the individuals to be present at the beginning of the observation period. Some of the individuals may join the study as time goes on. Let $z_i(t)\varepsilon\{1,\ldots,J\}$ denote the state in which individual i is at the beginning of time interval t. We define the indicator variables

$$Y_{ij}(t) = \begin{cases} 1 & \text{if individual } i \text{ is 'at risk' (under observation) at the beginning of time interval } t \text{ and } z_i(t)=j \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

The individuals with $Y_{ij}(t)=1$ constitute the risk sets

$$R_j(t) = \{i: Y_{ij}(t)=1\}, \qquad j=1,\ldots,J. \tag{2.2}$$

$R_j(t)$ contains all individuals who are at risk during time interval t and who are in state j. Note that an individual cannot belong to more than one risk set in time interval t. If the sample size n is fixed in advance, then $i=1,\ldots,n$. If new individuals join the study as time goes on, let n_t denote the number of individuals in the study up to time interval t. Then individuals $i=1,\dots,n_t$ will have to be investigated to determine the risk sets $R_j(t)$.

Suppose then, that for every individual i and time interval t such that $Y_{ij}(t)=1$ for some $j\epsilon\{1,\dots,J\}$, a p-vector $X_i(t)$ of relevant covariates is measured. The covariate process may include fixed or external time dependent as well as internal time dependent covariates (see Kalbfleisch and Prentice, 1980, ch.5.2). The vector of covariates may contain metric or dummy variables or both. As an approximation we assume that an individual can experience at most one event in time interval t, and if an individual is censored or is lost from the study, it is assumed that this happens at the end of the time interval. Similarly we think of the covariates as remaining fixed during each time interval, with the possible new value always determined at the beginning of the interval. When the interval lengths are small, this approximation is unlikely to influence statistical analysis a great deal.

Let $M(z_i(t))$ denote the set of attainable states from state $z_i(t)$ ordered in some way. Then we define the random variables $D_{ik}(t)$ as follows

$$D_{ik}(t) = \begin{cases} 1 \text{ if individual } i \text{ moves to state } k \text{ in time interval } t \\ 0 \text{ otherwise,} \qquad k\epsilon M(z_i(t)). \end{cases} \tag{2.3}$$

Furthermore, let $D_i(t)$ be the vector variable $\{D_{ik}(t)\colon k\epsilon M(z_i(t))\}$.

The history of the process up to time interval t is given by

$$F_t = \{(R_j(s),\ (X_i(s),\ D_i(s),\ i\epsilon R_j(s)),\ j=1,\dots,J),\ s\leq t\}$$

and

$$G_{tj} = F_{t-1} \cup \{R_j(t),\ (X_i(t),\ i\epsilon R_j(t))\},\ j=1,\dots,J;\ t=1,\dots,T, \tag{2.4}$$

F_t including and G_{tj} excluding the events which happen in time interval t. F_0 is assumed to represent initial information. If no initial information is available, we can take $F_0=\emptyset$.

The observation process is given by $\{(R_j(t),(D_i(t),X_i(t),i\epsilon R_j(t)),\ j=1,\dots,J),\ t=1,2,\dots\}$. Consider then a partially specified statistical model for the observation process. Especially we specify the conditional distributions of $D_i(t)$, given G_{tj}. It is assumed that the conditional distributions $P(D_i(t)|G_{tj})$ depend on the linear predictor $\beta^{t}_{jk}{}'X_i(t) + \theta_i$, where β^{t}_{jk} are parameter vectors common to all of the individuals and θ_i are individual specific parameters. The β^{t}_{jk} may depend on the origin state, the destination state, and the time interval t.

Several model specifications are possible. A dynamic form of the discrete-time logistic regression model is given by

$$P(D_i(t)|G_{tj}) = \frac{\prod_{k\epsilon M(j)} \exp(\beta^{t}_{jk}{}'X_i(t)+\theta_i)^{D_{ik}(t)}}{1 + \sum_{k\epsilon M(j)} \exp(\beta^{t}_{jk}{}'X_i(t) + \theta_i)} \; Y_{ij}(t) \; . \tag{2.5}$$

A multivariate probit specification for $P(D_i(t)|G_{tj})$ can also be used instead of (2.5).

Alternative specifications arise if the discreteness of the failure time data is due to the grouping of data from an underlying continuous distribution. One can start with the continuous-time hazard rate or distribution of failure times and then derive discrete-time hazard rates and distributions for grouped data. In general this involves integrals of the density function over the grouping intervals and computations may become laborious but in some cases derivation of the distribution for the grouped model is tractable. Consider for example a proportional hazards model where the transition specific hazard rate for an individual being in state j is given by

$$\lambda_{jk}(t|X_i) = \lambda_{oj}(t) \;\; \exp(\gamma_{jk}+\beta_{jk}'X_i+\theta_i), \;\; k\epsilon M(j)$$

(for uniqueness set one of the γ_{jk}'s equal to zero; the covariates are assumed to be time independent). Then it can be shown that the conditional probabilities $P(D_i(t)|G_{tj})$ are

$$P(D_i(t)|G_{tj}) = (1 - \lambda_{ojt}^{\sum_k c_{ijk}}) \frac{\prod_k c_{ijk}^{D_{ik}(t)}}{\sum_k c_{ijk}} \qquad \text{if } \sum_k D_{ik}(t) = 1, \text{ and}$$

$$P(D_i(t)|G_{tj}) = \lambda_{ojt}^{\sum_k c_{ijk}} \qquad \text{if } \sum_k D_{ik}(t) = 0 \tag{2.6}$$

where $c_{ijk} = \exp(\gamma_{jk}+\beta_{jk}'X_i+\theta_i)$. The λ_{ojt} in (2.6) are given by

$$\lambda_{ojt} = \exp(-\int_{a_{t-1}}^{a_t} \lambda_{oj}(u)du)$$

where a_{t-1} and a_t denote lower and upper bound of time interval t (a_o=0). (2.6) may be generalized to include time dependent covariates.

The individual specific parameters are included to account for the effect of unobserved variables (unobserved heterogeneity). A convenient approach is to assume a parametric distribution for the heterogeneity component θ_i (individual parameter) and to estimate the regression coefficients β together with the unknown parameters of the distribution of the heterogeneity component from a 'marginal' likelihood integrating out the unobserved heterogeneity component. Such a model is referred to as a random effect model. If the heterogeneity component is treated as a parameter then the model is referred to as a fixed effect model. The θ_i are incidental parameters (in the sense of Neyman and Scott (1948)) and β_{jk}^t, which is common to all individuals in the sample, is a vector of structural parameters. A basic statistical issue is to develop an estimator for β_{jk}^t that has good properties in this case. A suitable estimation procedure is presented in section 4 for the logistic model.

The use of a fixed effect model will be more appropriate if the individual effects and the included explanatory variables are correlated and if one is not able to give an exact specification of the conditional distribution of θ, given the explanatory variables. Treating the individual effect as an unknown parameter is equivalent to adding a time invariant variable to the set of explanatory variables. Therefore, using a fixed effect model can eliminate the bias arising from the correlation between the unobserved time invariant effects and the included explanatory variables, whereas a random effect inference ignoring the correlation between the effects and explanatory variables can lead to biased estimation. Furthermore, in the fixed effect approach there is no need to postulate a specific distribution of θ. Estimation procedures for the fixed effect approach as well as for the random effect approach are discussed in section 4.

3. Some special cases of interest

3.1 Repeated events of the same kind

This is a one-state process, for example, birth in a fertility history or the lifetimes of an electric appliance until the occurence of a certain defect or break-down. Here, indicator variables $Y_i(t)$ and $D_i(t)$ are defined as follows

$$Y_i(t) = \begin{cases} 1 & \text{if individual is at risk at the beginning of time interval } t \\ 0 & \text{otherwise ,} \end{cases}$$

and

$$D_i(t) = \begin{cases} 1 & \text{if individual } i \text{ experiences an event in time interval } t \\ 0 & \text{otherwise .} \end{cases}$$

The history of the process up to time t is defined analogously, dropping the subscript j in G_{tj} because there is only one state. The conditional probabilities $P(D_i(t)|G_t)$ are again assumed to be functions of a linear predictor $\beta_t{}'X_i(t)$ and individual specific parameters θ_i.

A probit specification of the conditional probabilities is

$$P(D_i(t)|G_t) = \Phi[(\beta_t{}'X_i(t) + \theta_i)(2D_i(t) - 1)]\; Y_i(t) \qquad (3.1)$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. A logistic regression model is given by

$$P(D_i(t)|G_t) = \frac{\exp(\beta_t{}'X_i(t) + \theta_i)^{D_i(t)}}{1 + \exp(\beta_t{}'X_i(t) + \theta_i)}\; Y_i(t) \; . \qquad (3.2)$$

Other specifications arising from the grouping of a continuous-time model can also be used.

3.2 A two-state model

If there are only two states z_1 and z_2 of interest, e.g. employed - unemployed, then we define the indicator variables $Y_i(t)$ as before and random variables $D_{i1}(t)$ and $D_{i2}(t)$ according to (2.3). The parameter vector β_2 describes the influence of the covariates or prognostic factors on the conditional transition probabilities $P(D_{i2}(t)|G_t)$ from state z_1 to state z_2, whereas the parameter vector β_1 represents the regression coefficients for the transition from state z_2 to state z_1.

Probit and logit specifications are given by

$$P(D_{i2}(t)|G_t,z_i(t)=z_1) = \Phi[(\beta_{2t}'X_i(t) + \theta_i)(2D_{i2}(t) - 1)]\ Y_i(t),$$

$$P(D_{i1}(t)|G_t,z_i(t)=z_2) = \Phi[(\beta_{1t}'X_i(t) + \theta_i)(2D_{i1}(t) - 1)]\ Y_i(t)$$

$$P(D_{i2}(t)|G_t,z_i(t)=z_1) = \frac{\exp(\beta_{2t}'X_i(t) + \theta_i)^{D_{i2}(t)}}{1 + \exp(\beta_{2t}'X_i(t) + \theta_i)}\ Y_i(t) ,$$

$$P(D_{i1}(t)|G_t,z_i(t)=z_2) = \frac{\exp(\beta_{1t}'X_i(t) + \theta_i)^{D_{i1}(t)}}{1 + \exp(\beta_{1t}'X_i(t) + \theta_i)}\ Y_i(t) .$$

3.3 Sojourn time in a given state

A special case which is important for practical situations arises if interest is restricted to a certain state and if we investigate the exit rate from this state. Here the end of the first episode is not necessarily the beginning of the second episode and the end of the second episode is usually not the beginning of the third, etc. For example, the successive employment spells of a person can be interrupted by unemployment, further education, illness, etc. In our general model we take this into account by restricting the risk set and the random variables $D_i(t)$ on the state under consideration. $Y_i(t)$ and $D_i(t)$ are defined as follows

$$Y_i(t) = \begin{cases} 1 & \text{if individual } i \text{ is at risk at the beginning of time interval } t \text{ and } z_i(t)=z \\ 0 & \text{otherwise ,} \end{cases}$$

$$D_i(t) = \begin{cases} 1 & \text{if individual } i \text{ leaves state } z \text{ in time interval } t \\ 0 & \text{otherwise .} \end{cases}$$

Specifications of the conditional transition probabilities $P(D_i(t)|G_t)$

are as described in (3.1) and (3.2).

4. Maximum likelihood estimation

The present section deals with the maximum likelihood estimation of the unknown parameters based on the general model derived in section 2. First we evaluate a general expression for the likelihood of the observation process which corresponds to data collected up to time interval T. In order to keep such a likelihood expression in a manageable form we restrict the way in which the law of the process is allowed to depend on the parameters β and θ. The assumptions generalize those of Arjas (1984). The resulting likelihood function represents a joint likelihood function for the structural parameters and the individual specific parameters as well. One disadvantage is that structural and individual specific parameters cannot be estimated consistently from the joint likelihood function if the number of time intervals is small. Therefore, our next step is to derive a conditional likelihood given a suitable sufficient statistic for the individual specific parameter. This conditional likelihood does not depend on the individual parameters and the structural parameters can be estimated by maximizing the conditional likelihood function. However, the conditional approach only applies to the logistic representation of the conditional transition probabilities. In the last section we investigate the random effect approach which is applicable for all specifications of the conditional transition probabilities if the distribution of the individual specific parameters is known.

From the resulting likelihood expression it becomes clear that the estimation procedures are also applicable if some of the first episodes are left-censored.

4.1 The joint likelihood function

The observation process for the general model of section 2 is

$$\{(R_j(t),\ (D_i(t),\ X_i(t),\ i\in R_j(t)),\ j=1,..,J),\ t=1,...,T\}\ ,$$

corresponding to data collected up to time T, and the likelihood is the joint probability of the observation process. Using some properties of conditional probabilities it can be shown that

$$L(\beta,\theta) = \prod_{t\leq T} P((R_j(t),\ (D_i(t),\ X_i(t),\ i\epsilon R_j(t)),\ j=1,\ldots,J)|F_{t-1}). \quad (4.1)$$

Now we impose the following assumption.

Assumption 1

For each t and (β,θ) the random variables $(R_j(t),\ (D_i(t),\ X_i(t),\ i\epsilon R_j(t)),$ $j=1,\ldots,J$, are conditionally independent, given F_{t-1}.

The assumption concerns the conditional independence between the risk sets respectively the individuals who constitute the risk sets. This assumption is likely to hold in practice. Then the likelihood function is given by

$$L(\beta,\theta) = \prod_{t\leq T} \prod_{j=1}^{J} P(R_j(t),\ (D_i(t),\ X_i(t),\ i\epsilon R_j(t))|F_{t-1})$$

$$= \prod_{t\leq T} \prod_{j} P(D_i(t),\ i\epsilon R_j(t)|R_j(t),\ (X_i(t),\ i\epsilon R_j(t)),\ F_{t-1})$$

$$P(R_j(t),\ (X_i(t),\ i\epsilon R_j(t))|F_{t-1}) \quad (4.2)$$

The second term on the right hand side of (4.2) is the joint probability of $R_j(t)$, the individuals at risk in state j during time interval t, and the covariates $\{X_i(t),\ i\epsilon R_j(t)\}$ measured for these individuals, conditional on the history F_{t-1}. In the following we assume that this probability, given F_{t-1}, does not depend on β and θ.

Assumption 2

For each t, the conditional distribution of $(R_j(t),\ (X_i(t),\ i\epsilon R_j(t)),$ $j=1,\ldots,J$, given F_{t-1}, does not depend on β and θ.

The assumption states that, given the knowledge contained in F_{t-1}, knowing also the values of $R_j(t)$ and $\{X_i(t),\ i\epsilon R_j(t)\}$ does not contain additional information about β and θ. Note that in the case where the random variables $R_j(t)$ govern the right censoring of the individuals the assumption implies that such censoring is noninformative (see Kalbfleisch and Prentice (1980), ch. 5.2). For a further discussion of the assumptions see Arjas (1984).

If the assumption does not hold, the likelihood expressions mentioned below can be considered as partial likelihood functions (see Cox (1975)). Otherwise, it becomes necessary to specify the conditional probabilities in the second term on the right hand side of (4.2).

In addition we impose a third assumption which is strictly connected to assumption 1.

Assumption 3

For each risk set $R_j(t)$, $j=1,\dots,J$, $t=1,\dots,T$, and for each (β,θ) the random variables $D_i(t)$, $i\epsilon R_j(t)$, are conditionally independent, given G_{tj}.

Assumption 3 again states a conditional independence assumption between the individuals in the sample.

Then the relevant factor of the likelihood which is again denoted by $L(\beta,\theta)$ becomes proportional to the expression

$$L(\beta,\theta) = \prod_{t\leq T} \prod_{j} \prod_{i\epsilon R_j(t)} P(D_i(t)|G_{tj}) \tag{4.3}$$

The conditional probabilities on the right hand side of (4.3) have still to be specified. For this purpose we can use one of the models discussed in the previous sections.

But if we use (4.3) as a joint likelihood function for the parameters β and θ, a difficulty arises. The parameters β and θ cannot be estimated consistently from this joint likelihood if the number T of time intervals is finite. The reason is that the number of individual specific parameters increases with sample size. Andersen (1973, p. 68-71) considers the binary logit model with T=2 and one structural parameter β. He shows that plim $\hat{\beta}$ = 2β. The same result is obtained for any symmetric distribution, not just the logistic one. Heckman (1981b, p. 187) gives an heuristic argument. He points out that the roots of the likelihood equations involve the joint solution of structural and individual specific parameters. Since estimators of θ_i are necessarily inconsistent, if T is finite, the inconsistency of the estimator for the individual specific parameters is then transmitted to the estimator for the structural parameters.

The inconsistency decreases if T becomes large and in the limit ($T\to\infty$) disappears. Therefore, estimation of β (and θ) by maximizing the joint likelihood function (4.3) can be used, if the number of time intervals is moderate or large. But further Monte Carlo studies are needed to determine the size of T such that the maximization of the joint likelihood function performs satisfactory estimates.

4.2 A conditional likelihood function

Our next step is to derive an alternative approach using a conditional likelihood function. The key idea is to base the likelihood function on the conditional distribution of the data, conditioning on a set of sufficient statistics for the individual parameters. But this approach only applies to the logistic model. For this model a sufficient statistic for the individual specific parameter is given by

$$N_i = \sum_{t=1}^{t_i} \sum_k D_{ik}(t) \tag{4.4}$$

where t_i=max{t: $Y_{ij}(t)=1$ for some $j\epsilon\{1,\ldots,J\}$}. N_i is the number of completed spells of individual i.

We consider the conditional probability

$$P((R_j(t),\ (D_i(t),\ X_i(t),\ i\epsilon R_j(t)),\ j=1,\ldots,J),\ t=1,\ldots,T \mid N_i,\ i=1,\ldots,n) \tag{4.5}$$

Since the event counts $D_i(t)$ are part of the event which defines the condition in (4.5), we rewrite (4.5) as the quotient

$$\frac{P((R_j(t),\ (D_i(t),\ X_i(t),\ i\epsilon R_j(t)),\ j=1,\ldots,J),\ t=1,\ldots,T)}{P(N_1,\ldots,N_n)}. \tag{4.6}$$

The probability in the nominator of (4.6) is given by (4.3) multiplied by a factor which does not depend on the parameters because of assumptions 1 to 3. Substitution of the logistic representation (2.5) for the conditional probabilities $P(D_i(t)|G_{tj})$ into (4.3) yields

$$L(\beta,\theta) = \prod_{t\leq T} \prod_j \prod_{i=1}^{n} \frac{\exp(\sum_k \beta_{jk}^{t}{}'X_i(t)\ D_{ik}(t)Y_{ij}(t) + \theta_i \sum_k D_{ik}(t)Y_{ij}(t))}{(1 + \sum_{k\epsilon M(j)} \exp(\beta_{jk}^{t}{}'X_i(t) + \theta_i))^{Y_{ij}(t)}}$$

$$= \prod_{i=1}^{n} \frac{\exp(\sum_t \sum_j \sum_k \beta_{jk}^{t}{}'X_i(t)Y_{ij}(t)\ D_{ik}(t) + \theta_i \sum_t \sum_j \sum_k \Sigma D_{ik}(t)Y_{ij}(t))}{\prod_t \prod_j (1 + \sum_{k\epsilon M(j)} \exp(\beta_{jk}^{t}{}'X_i(t) + \theta_i))^{Y_{ij}(t)}} \tag{4.7}$$

The probability in the denominator of (4.6) is given by

$$P(N_1,\dots,N_n) = \prod_{i=1}^{n} \sum_{\sum_{t=1}^{t_i}\sum_k D_{ik}(t)=N_i} \frac{\exp(\sum_t \sum_j \sum_k \beta_{jk}^{t}{}'X_i(t)Y_{ij}(t)D_{ik}(t) + \theta_i N_i)}{\prod_t \prod_j (1 + \sum_{k\epsilon M(j)} \exp(\beta_{jk}^{t}{}'X_i(t) + \theta_i))^{Y_{ij}(t)}} \tag{4.8}$$

and the conditional likelihood function which is denoted by CL(β) is obtained by dividing (4.7) by (4.8)

$$CL(\beta) = \prod_{i=1}^{n} \frac{\exp(\sum_t \sum_j \sum_{k\epsilon M(j)} \beta_{jk}^{t}{}'X_i(t)Y_{ij}(t)D_{ik}(t))}{\sum_{\sum_t\sum_k D_{ik}(t)=N_i} \exp(\sum_t \sum_j \sum_{k\epsilon M(j)} \beta_{jk}^{t}{}'X_i(t)Y_{ij}(t)D_{ik}(t))} \tag{4.9}$$

The conditional likelihood function (4.9) only depends on the structural parameters β, and does not depend upon the individual specific parameters. Hence standard asymptotic theory applies. The conditional ML-estimator of β is consistent and asymptotically normally distributed provided that the individual parameters and the conditional likelihood function satisfy regularity conditions (see Andersen 1973).

Note that individuals with $N_i=0$ or $N_i=t_i$ where $t_i=\max\{t: Y_{ij}(t)=1$ for some $j\epsilon\{1,\dots,J\}\}$ do not contribute any information to the conditional likelihood (4.9), since for these values of N_i nominator and denominator on the right hand side of equation (4.9) are equal. Therefore, the number of individuals who have one or more completed spells should be at least moderate.

We must keep in mind that the conditional likelihood method is only helpful in a logit model. It is not generally possible to find minimum sufficient statistics for the individual specific parameters which are independent of the structural parameters and which have a smaller dimension than the sample size. This is possible if the distribution is a member of the exponential family like the logistic parametrization of the multinomial distribution. Therefore, conditional likelihood methods are not a general approach in fixed effect models, but if the logistic representation is appropriate, the conditional approach has some advantages. It does not require a specification for the distribution of θ. If one makes such an assumption, the distribution of the heterogeneity θ conditional on the observed covariates X is needed, and in general this should be allowed to depend upon the observed covariates. If there was omitted variable bias before introducing θ, and if one mistakenly models θ as independent of X, then the resulting estimator based on the 'marginal' likelihood (see next section) will also be biased. The fixed effect approach presented here has the advantage of allowing for a very general relationship between θ and X.

Note that the appropriateness of the logistic representation can in principle be tested by one of the specification tests for the multinomial logit model described by Hausman and McFadden (1984) in a choice theoretic context.

4.3 A marginal likelihood function

In this section we discuss an alternative approach assuming that the individual specific parameters follow a distribution. The individual specific parameter θ, the heterogeneity, is not observable. Let $G(\theta)$ denote the (marginal) distribution of θ. In this case the probabilities $P(D_i(t)|G_{tj},\theta_i)$ are conditional probabilities, given G_{tj} and θ_i, and the resulting likelihood function (4.3) is also conditional on the individual specific components θ_i. Introducing the indicator variables $Y_{ij}(t)$ as defined in (2.1) we can rewrite the contribution of individual i to the likelihood expression in (4.3)

$$L_i(\beta|\theta_i) = \prod_{t=1}^{T} \prod_j P(D_i(t)|G_{tj},\theta_i)^{Y_{ij}(t)} . \tag{4.10}$$

If it is possible to specify $G(\theta)$ as a member of a parametric class of probability distributions, estimation of the structural parameters can be based on the marginal distribution of the observation process integrating out the individual component θ. The marginal likelihood function is denoted by $ML(\beta,\gamma)$ where γ is the parameter vector determining the distribution of θ. It is given by

$$ML(\beta,\gamma) = \prod_{i=1}^{n} \int \prod_{t=1}^{T} \prod_j P(D_i(t)|G_{tj},\theta)^{Y_{ij}(t)} dG(\theta) . \tag{4.11}$$

The marginal likelihood function is a function of β and the unknown parameters γ of the population distribution $G(\theta)$. Maximization of this likelihood function will, under weak regularity conditions, give consistent and asymptotically normally distributed estimators for β and γ.

Note that in this approach the population distribution $G(\theta)$ is assumed to be known except for a finite number of parameters. Furthermore, if θ_i and $X_i(t)$ are correlated, we have to specify the joint distribution of (θ_i, X_i') in order to obtain consistent estimates of structural parameters. A convenient possibility in analogy to the linear model case is to assume that the dependence is only via a linear regression function (Chamberlain (1980, 1984))

$$\theta_i = \pi' X_i + \varepsilon_i \quad , \quad i=1,\ldots,n \tag{4.12}$$

where $X_i' = (X_i(1)', \ldots, X(t_i)')$, t_i as defined in (4.4), and where ε_i is independent of X_i. We assume that the ε_i are independent and identically distributed with distribution function $H(\varepsilon)$. Substitution of (4.12) into (4.10) and (4.11) yields a marginal likelihood function which is appropriate if the heterogeneity component is correlated with the observed covariates. This seems to be rather the rule than the exception.

For illustration let us consider the special case described in section 3.1 with the probit specification (3.1). Then, allowing for correlation between θ_i and X_i and using (4.12) the marginal likelihood function is given by (the parameter vector determining $H(\varepsilon)$ is denoted by α)

$$ML(\beta,\alpha) = \prod_{i=1}^{n} \int \prod_{t=1}^{T} \prod_{j} \{\Phi((\beta_t' X_i(t) + \pi' X_i + \varepsilon)(2D_i(t)-1))\}^{Y_{ij}(t)} dH(\varepsilon) \tag{4.13}$$

Note that sometimes identification problems may arise especially if the parameters β are assumed to be time independent.

Finally we mention alternative approaches of Liang and Zeger (1986) and Stiratelli, Laird and Ware (1984).

Liang and Zeger propose methods for longitudinal data (not for event history or failure time data) only assuming a functional form for the marginal distribution at each time corresponding to $P(D_i(t))$ in the present paper. The marginal distribution is assumed to belong to the family of generalized linear models. In addition, a covariance structure for $(D_i(1), \ldots, D_i(T))$ is assumed but this covariance structure across time is treated as a nuisance. Then they derive estimating equations similar to the quasi-likelihood approach (see, for example, McCullagh (1983)) and investigate asymptotic properties of the estimators of the regression coefficients.

Stiratelli, Laird and Ware (1984) consider the special case of longitudinal data with binary outcomes. They split up the set of covariates into two sets. The first set contains the covariates which vary over time and in the second set are the covariates which are fixed. The fixed covariates are denoted by x_i and the time varying covariates are denoted by z_{it}. Furthermore, let λ_i denote the T-vector of logits for individual i. Then, Stiratelli, Laird and Ware (1984) investigate a two-stage approach where at stage 1 they let

$$\lambda_i = X_i\beta + Z_i\alpha_i$$

with suitable defined matrices X_i and Z_i, and at stage 2 they assume that α_i is multivariate normal with expectation 0 and covariance matrix Σ. These assumptions define a general mixed model for the logits of the response probabilities and one could try to carry over this model into the event history or failure time context using the EM algorithm or empirical Bayes strategies for estimation.

References

Andersen EB. (1973) Conditional inference and models for measuring. Kopenhagen

Andersen PK, Gill RD (1982) Cox's regression model for counting processes: a large sample study. Annals of Statistics 10:1100-1120

Allison P (1982) Discrete-time methods for the analysis of event histories. In: Leinhardt S (ed) Sociological Methodology, pp. 61-98

Arjas E (1984) Logistic regression in modelling hazard: a fully parametric approach. Preprint, University of Oulu, Finland

Arjas E, Haara P (1986) A logistic regression model for hazards: asymptotic results. Scandinavian Journal of Statistics 13 (forthcoming)

Blossfeld HP, Hamerle A, Mayer KU (1986) Ereignisanalyse: Statistische Theorie und Anwendungen in den Wirtschafts- und Sozialwissenschaften. Campus, Frankfurt

Borgan Ø (1984) Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. Scandinavian Journal of Statistics 11:1-16, Correction 11:275

Breslow NE (1974) Covariance analysis of censored survival data. Biometrics 30:89-99

Chamberlain G (1980) Analysis of covariance with qualitative data. Review of Economic Studies 17:225-238

Chamberlain G (1984) Panel data. Handbook of Econometrics, Vol. II (Z. Griliches & M Intriligator, Eds.), North-Holland, Amsterdam

Chamberlain G (1985) Heterogeneity, omitted variable bias, and duration dependence. In: Heckman J, Singer B (eds) Longitudinal analysis of labor market data. Cambridge University Press, Cambridge

Cox DR (1975) Partial likelihood. Biometrika 62:269-276

Cox DR, Oakes D (1984) Analysis of survival data. Chapman and Hall, London

Haberman SJ (1977) Maximum likelihood estimates in exponential response models. Annals of Statistics 5:815-841

Hamerle A (1984) Zur statistischen Analyse von Zeitverläufen. Diskussionsbeitrag Nr 180, Universität Regensburg

Hamerle A (1985) Regressionsmodelle für gruppierte Verweildauern und Lebenszeiten. Zeitschrift Operations Research 29, B:243-260

Hanefeld U (1984) Das Sozio-ökonomische Panel - eine Längsschnittstudie für die Bundesrepublik Deutschland. Vierteljahreshefte zur Wirtschaftsforschung 4/84:391-406

Hausman J, McFadden D (1984) Specification tests for the multinomial logit model. Econometrica 52:1219-1240

Heckman J (1981a) Statistical models for discrete panel data. In: Manski CF, McFadden D (eds) Structural analysis of discrete data with econometric applications, MIT Press, Cambridge, Mass.:114-178

Heckman J (1981b) The incidental parameter problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski CF, McFadden D (eds) Structural analysis of discrete data with econometric applications. MIT Press, Cambridge, Mass.:179-195

Heckman J, Singer B (1984) Econometric duration analysis. Journal of Econometrics 24:63-132

Hujer R, Schneider H (1986) Ökonometrische Ansätze zur Analyse von Paneldaten: Schätzung und Vergleich von Übergangsratenmodellen. Manuskript, Frankfurt

Kalbfleisch JD, Prentice RL (1980) The statistical analysis of failure time data. Wiley, New York

Laird N, Olivier D (1981) Covariance analysis of censored survival data using log-linear analysis techniques. Journal of the American Statistical Association 76:231-240

Lawless JF (1982) Statistical models and methods for lifetime data. Wiley, New York

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13-22

Mantel N, Hankey BF (1978) A logistic regression analysis of response-time data where the hazard function is time dependent. Communications in Statistics A7:333-347

McCullagh P (1983) Quasi-likelihood functions. Annals of Statistics 11:59-67

Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. Econometrica 16:1-32

Prentice RL, Gloeckler LA (1978) Regression analysis of grouped survival data with applications to breast cancer data. Biometrics 34:57-67

Stiratelli R, Laird N, Ware J (1984) Random effect models for serial observations with binary responses. Biometrics 40:961-971

Thompson WA Jr (1977) On the treatment of grouped observations in life studies. Biometrics 33:463-470

Tuma NB, Hannan MT (1984) Social dynamics: Models and methods. Academic Press, New York

Prof. Dr. A. Hamerle
Fakultät für Wirtschafts-
wissenschaften und Statistik
Universität Konstanz
D-7750 Konstanz